

THEORY AND METHODS

Criteria for evaluating evidence on public health interventions

L Rychetnik, M Frommer, P Hawe, A Shiell

J Epidemiol Community Health 2002;56:119–127

See end of article for authors' affiliations

Correspondence to: Lucie Rychetnik, Effective Healthcare Australia, Victor Coppleson Building, DO2, University of Sydney, NSW 2006, Australia; lucier@med.usyd.edu.au

Accepted for publication 30 July 2001

Public health interventions tend to be complex, programmatic, and context dependent. The evidence for their effectiveness must be sufficiently comprehensive to encompass that complexity. This paper asks whether and to what extent evaluative research on public health interventions can be adequately appraised by applying well established criteria for judging the quality of evidence in clinical practice. It is adduced that these criteria are useful in evaluating some aspects of evidence. However, there are other important aspects of evidence on public health interventions that are not covered by the established criteria. The evaluation of evidence must distinguish between the fidelity of the evaluation process in detecting the success or failure of an intervention, and the success or failure of the intervention itself. Moreover, if an intervention is unsuccessful, the evidence should help to determine whether the intervention was inherently faulty (that is, failure of intervention concept or theory), or just badly delivered (failure of implementation). Furthermore, proper interpretation of the evidence depends upon the availability of descriptive information on the intervention and its context, so that the transferability of the evidence can be determined. Study design alone is an inadequate marker of evidence quality in public health intervention evaluation.

Appraisal of evaluative research used in evidence-based health care centres on three major questions. Firstly, is the research good enough to support a decision on whether or not to implement an intervention? Secondly, what are the research outcomes? Thirdly, is the research transferable to the potential recipients of the intervention (individuals or populations)?¹

In this paper we ask whether (or to what extent) evaluative research on *public health interventions* can be adequately appraised by applying well established criteria for appraising evidence about prevention and treatment in *clinical practice*.^{2–5} We adduce that these criteria are very useful in evaluating some important aspects of evidence. However, there are other important aspects of evidence relevant to public health interventions that are not covered by the established criteria. We draw attention to these additional aspects of evidence and explain their importance in the assessment of public health interventions. We emphasise the distinction between the appraisal of evidence and the process of making policy or operational decisions on the implementation of interventions. Research-based evidence is only one of several factors to be taken into account in these decisions.

Public health interventions tend to be complex, programmatic, and context dependent. The evidence for their effectiveness must be sufficiently comprehensive to encompass that complexity. The evaluation of evidence must distinguish between the fidelity of the evaluation process in detecting the success or failure of an intervention, and the relative success or failure of the intervention itself. Moreover, if an intervention is unsuccessful, the evidence should help to determine whether the intervention was inherently faulty (that is, failure of intervention concept or theory), or badly delivered (failure of implementation).⁶ Furthermore, proper interpretation of the evidence depends upon the availability of adequate descriptive information on the intervention and its context, so that the transferability of the evidence can be determined.

To fulfil these requirements, we suggest an expansion of the criteria that are used in clinical medicine for appraising research. We draw on evidence-evaluation schema that were

developed for epidemiological and qualitative research, health promotion programme evaluations and health economic evaluations.

DEFINITIONS

For the purposes of this paper, an *intervention* is defined as a set of actions with a coherent objective to bring about change or produce identifiable outcomes. These actions may include policy, regulatory initiatives, single strategy projects or multi-component programmes. *Public health interventions* are intended to promote or protect health or prevent ill health in communities or populations. They are distinguished from clinical interventions, which are intended to prevent or treat illness in individuals. *Context* refers to the social, political and/or organisational setting in which an intervention was evaluated, or in which it is to be implemented. The contextual characteristics that are relevant vary with the type of intervention. Important contextual characteristics for a public health intervention might include factors in the political and organisational environment and socioeconomic or demographic features of the population.

Evaluation is a process of determining the value or worth of something by judging it against explicit, predetermined standards.⁷ *Evidence* comprises the interpretation of empirical data derived from formal research or systematic investigations, using any type of science or social science methods. This definition of evidence is purposefully circumscribed to articulate the scope of this paper. In our consideration of *evidence-based practice*, we focus on evidence about likely consequences of interventions, such as effectiveness and cost effectiveness, not evidence about need for services. Thus, we distinguish between data on the cause or scale of a health problem (aetiological studies and needs assessment) and evidence on the implementation and outcomes of interventions. This paper deals with the latter.

Until recently public health epidemiology was chiefly concerned with aetiological hypotheses, rather than evaluative hypotheses. Intervention evaluation has its origins in the social sciences, notably education and psychology.^{8,9} To

Table 1 Schemata for appraising quantitative evaluations of intervention effectiveness

Type of schema and evidence to be appraised	How the schema works and the criteria used
<p><i>Critical appraisal checklists for quantitative studies of intervention effectiveness</i></p> <p>Checklists derived from the evidence based medicine working group (EBMWWG), which appraise articles about clinical therapy or prevention.^{4,5,14}</p>	<p>Guide appraisal of the validity and applicability of published evidence. Structures critical appraisal into 3 sections:</p> <ul style="list-style-type: none"> • Are the results valid? • What were the results? • Will the results help in caring for patients? <p>The validity of evaluative research is judged on the level of evidence (study design and its potential for eliminating bias, e.g. systematic review of RCTs is the highest level of evidence); and the implementation of methods and analysis.</p> <p>Clinical importance and applicability of the findings are determined by the magnitude (with confidence intervals) of the estimate of effect and relevance of the outcomes measured.</p>
<p>Checklist from Oxford based Public Health Resource Unit for appraising articles about health interventions.¹⁵</p>	<p>Focus on the validity of research, which is assessed on similar criteria to above: study design; selection bias; confounding; blinding; data collection and classification of outcomes; follow up, withdrawal and drop out and analysis.</p>
<p>Critical appraisal checklist for evaluating research on public health interventions, from the Effective Public Health Practice Project, Ontario Ministry of Health.¹⁶</p>	<p>Ontario checklist has similar criteria as above, but also considers the integrity of intervention being evaluated.</p>
<p><i>Critical appraisal checklists within a guide for preparing systematic reviews.</i></p> <p>Critical appraisal criteria in Cochrane Collaboration Handbook for Reviewers in the Cochrane Library. Schema is part of a guide for preparing and disseminating systematic reviews of RCTs on intervention effectiveness.¹⁷</p>	<p>The quality of an RCT is assessed on criteria as above:</p> <ul style="list-style-type: none"> • Assignment to treatment and control groups and blinding • Degree of potential confounding • The classification of outcomes and follow up • Appropriate analysis, for example, analysis by 'intention to treat'
<p>The Campbell Collaboration was established for preparing systematic reviews on social and education interventions.¹⁸</p>	<p>Approaches to evaluating evidence are under consideration, discussion papers available on their website.</p>
<p><i>Guides for preparing or evaluating reviews and clinical guidelines</i></p> <p>A guide for evaluating reviews, RCTs and non-randomised observation studies: Method for Evaluating Research and Guideline Evidence (MERGE).¹⁹</p>	<p>Critical appraisal checklists for reviews, RCT and observation studies (non-randomised controlled trial, cohort, case-control, before and after and interrupted time series). Evaluation criteria are grouped according to:</p> <ul style="list-style-type: none"> • Descriptive information about review or study (for example, type of intervention, implementation, outcomes considered, potential confounders and characteristics of population and setting) • Study design, implementation and analysis • Overall assessment of credibility of findings
<p>A handbook from the Australian NH and MRC on how to assess and apply research evidence. Part of a series on preparing practice guidelines.²⁰</p>	<p>Structures critical appraisal into three questions regarding the evidence:</p> <ul style="list-style-type: none"> • Is there a real effect? (Strength of evidence: level, quality and statistical precision) • Is the size of the effect clinically important? (Size of effect) • Is the evidence relevant to practice? (Relevance of evidence)
<p><i>Rules of evidence used to formulate graded recommendations for action</i></p> <p>Rules of evidence from the Canadian Task Force on the Periodic Health Examination, which made recommendations on preventive interventions in primary settings.^{2,11}</p>	<p>Recommendations for action are determined by a systematic review of studies of effectiveness, which include consideration of:</p> <ul style="list-style-type: none"> • Level of evidence • Quality of study methods • Number of studies • Magnitude of effect • Consistency of findings • Generalisability of findings to primary care setting
<p>Criteria from the US Preventive Services Task Force, which also formed recommendations on clinical preventive interventions.³</p>	<p>Strength, Class or Grade of Recommendations primarily based on the level of evidence (study design): Level I evidence is a systematic review of RCTs. Intervention cost and burden of disease are included if evidence on effectiveness is uncertain. Cost effectiveness to be potentially included future formulations of recommendations.</p>
<p>Rules of evidence from the Task Force on Community Preventive Services, which is forming recommendations about public health interventions.²¹</p>	<p>The Grade of recommendations made is linked to level of evidence, which determined by the study design. Levels of evidence in descending order are:</p> <ul style="list-style-type: none"> • Systematic review of RCTs with homogeneity (Level 1) • Individual RCT with narrow confidence interval • Systematic review of cohort studies, single cohort or RCT with <80% follow up • Systematic review of case-controls, individual case-control • Case series or poor quality cohort or case-control • Expert opinion without explicit critical appraisal or based on physiology, bench research or 'first principles'.
<p>The Oxford Centre for EBM have used Levels of Evidence to grade recommendations on therapy, prevention, aetiology and harm.²²</p>	<p>The likelihood of a causal relationship is determined by:</p> <ul style="list-style-type: none"> • Strength: magnitude of measured association • Consistency: repeated in multiple observations • Temporality: cause precedes effect • Biological gradient: a dose response relation • Coherence: no conflict with current knowledge • Plausibility: biological or theoretical • Experimental evidence: association examined using manipulation and controls (now considered a 'gold standard' demonstration of causality)
<p><i>Approaches to assessing causal associations and causal inference</i></p> <p>The schemata guide the appraisal of epidemiological evidence on causality (causal relations between two variables).^{23,34} The criteria can be applied to appraise evaluation research (evidence on the causal link between an intervention and its effects).</p>	<p>The likelihood of a causal relationship is determined by:</p> <ul style="list-style-type: none"> • Strength: magnitude of measured association • Consistency: repeated in multiple observations • Temporality: cause precedes effect • Biological gradient: a dose response relation • Coherence: no conflict with current knowledge • Plausibility: biological or theoretical • Experimental evidence: association examined using manipulation and controls (now considered a 'gold standard' demonstration of causality)

Table 1 continued

International Agency for Research on Cancer, (WHO) uses standards of evidence to appraise human and animal studies of carcinogenicity. ²⁵	Criteria for causation are used to appraise evidence when conducting an assessment of cancer risk. Weight of supporting evidence (causality) determined by: <ul style="list-style-type: none"> ● Study designs (epidemiological studies) ● Quality of studies ● Other studies (for example, animal studies) ● Plausible inferences about mechanism of action ● Other causal criteria (strong association, replication in multiple studies and consistency of findings)
<i>Quality assessment criteria for programme evaluations</i>	Examines the limitation of the intervention and the overall quality of the evaluation conducted. Considers: <ul style="list-style-type: none"> ● Stage of intervention ● Nature of objectives ● Target group specification ● Variables measures ● Instruments used to assess outcomes ● Evaluation study design
Guides the appraisal of health promotion programme evaluations. ²⁶	

strengthen the criteria for appraising evaluative research in public health we have drawn upon a broad-based literature beyond the fields of epidemiology and evidence-based medicine. We acknowledge, however, the limitation of relying on the English language literature.

IS THE RESEARCH GOOD ENOUGH?

If the research is good enough, it will confirm and quantify the causal relation between the intervention and its effects where such a relation exists. Good research will also help us to understand *why* an intervention appears to be effective or ineffective

Levels of evidence and causality

The assessment of causality for evidence-based health care has mostly depended upon the *level* of evidence, which traditionally has been defined by the study design used in evaluative research. Study designs are graded by their potential to eliminate bias. A hierarchy of study designs was first suggested by Campbell and Stanley in 1963,¹⁰ and levels of evidence based on study design were proposed by Fletcher and Sackett for the Canadian Taskforce on the Periodic Health Examination in 1979.¹¹ Systematic reviews of randomised controlled trials (RCTs) have become widely accepted as providing the best evidence (level 1) on the effects of preventive, therapeutic, rehabilitative, educational or administrative interventions in medicine.¹² The concept of levels of evidence has been widely adopted to determine the grade of recommendations for clinical practice, for example, in the recommendations of the US Preventive Services Task Force and the Canadian Task Force on the Periodic Health Examination.^{2,3} Levels of evidence have also been applied to other areas of evidence-based decision making in health, including prognosis, diagnosis and economic analysis.¹³

We have collated examples of existing guides for appraising evidence (table 1). Most of these guides are designed to help the user in assessing the factors that determine the existence and strength of a causal relation.

The guides differ in their scope. We have grouped them in table 1 according to their overall aim (left column), adding a summary of how each works and listing some of the criteria used (right column). The examples include critical appraisal checklists for quantitative studies of intervention effectiveness; guides on evaluating reviews or clinical guidelines and rules of evidence to formulate graded recommendations for action. It is standard practice in these guides to define the level of evidence in terms of the study design and to treat this as the primary determinant of credibility. Also included in table 1 are generic guides for determining causal inference in epidemiological research, which encompass criteria that can be applied to appraise causal relations in evaluation research.

Levels of evidence and public health interventions

The assessment of causality for public health interventions has also mostly depended upon the level of evidence.^{27–29} However, there is persisting controversy about the reliance on the study design as the main criterion of the credibility of evidence. The debate concentrates on the primacy of the RCT for evaluating public health interventions, with respect to (a) the difficulty of conducting RCTs for complex programmatic interventions, (b) the difficulty of interpreting their results, and (c) the tendency to downgrade the contribution of observational studies.

(a) RCTs and complex interventions

Many public health interventions require multiple, flexible and community driven strategies.^{30–32} RCTs have been described as unable to accommodate the complexity and flexibility that characterises such programmes. They are perceived as being feasible only for evaluating relatively simple, standardised and unvarying interventions and thus as being too rigid and inappropriate for public health settings.^{33,34}

Such criticisms of the RCT are based on a consideration of “classic” RCTs in which the intervention is standardised and the individual is the unit of randomisation. Cluster trials can accommodate communities, schools or other “clusters” as the unit of analysis, and RCTs can cope with non-standard interventions; points that seem to be lost by some trial critics.³⁴ RCTs have a long history of successful application in evaluating the effectiveness of social interventions.³⁵ Given the strength of this study design, the use of a non-randomised study in settings where RCTs would have been feasible represents a lost opportunity.³⁶ Our concern is that evaluators around the world may move away from favouring RCTs in public health for what we see as the wrong reasons; that is, a mistaken belief that experimental designs are only useful for evaluating standard, simple interventions aimed at individuals.

We reaffirm that a well conducted RCT is the best (albeit sometimes impractical) study design for determining a causal relation between an intervention and its putative outcomes. However, study design alone cannot suffice as the main criterion for the credibility of evidence about public health interventions.

(b) Interpretation of study results

Deficient a priori criteria for the adequacy of evidence on public health interventions have led to disagreements about interpretation of results, particularly negative findings.^{37,38} Some current appraisals of evidence do not assist in making a distinction between *failure to demonstrate underlying effectiveness* and *good evidence of ineffectiveness*.

Negative findings warrant careful exploration. Has the research failed to find an effect where one exists (evaluation failure)? Or is there truly no effect (programme failure)?²⁶ In the event of programme failure, is the failure attributable to an inherent inadequacy in the intervention (that is, a failure of intervention theory), or attributable to poor implementation? The authors of some systematic reviews have acknowledged that crucial factors such as the stability of the programme being evaluated, the quality of the implementation, or the adequacy of the outcome measures relative to programme goals, were not taken into account.³⁹ Without this information one cannot conclude that negative results mean that an intervention is ineffective. However, evidence of adequate implementation, and other measures to monitor the evaluation process, are important regardless of whether the findings are negative or positive.

The disagreements about interpretation of the results of community-based trials have been the basis for recommendations to expand the scope of evaluation methods for community programmes.⁴⁰⁻⁴³ These recommendations need to be supported by parallel developments in the criteria used to appraise the quality of evidence on public health interventions.

It has been proposed that evaluation designs should be more prudently and strategically sequenced to a programme's stage of development and to the available evaluation resources.^{6 44-47} Expensive randomised trial designs should be used only after research using simpler and cheaper designs has yielded satisfactory results regarding the feasibility of the intervention. Thus an RCT design may be best used to test a causal hypothesis, after satisfactory pre-post single group design has been conducted, and assurance has been obtained that the measuring instruments satisfactorily capture programme implementation processes and outcomes.⁴⁴

Specification of the theoretical basis of the intervention can also improve the credibility of outcome measures, and accords with a trend towards making the hypotheses and assumptions underpinning public health interventions more explicit.^{48 49} Intervention theories should be explicit and plausible. Explicit theories allow us to determine whether they are commensurable with the impact and outcome measures that have been adopted to evaluate that intervention, and whether an appropriate method was used to analyse those measures.⁵⁰ The trend towards identifying the anticipated causal pathway of an intervention (the "mode of action") is redressing the pragmatic "black box" use of epidemiology that placed more weight on research methods and outcomes than on intervention theory.^{51 52}

Multi-dimensional approaches are available for evaluating outcomes research.⁵³ Table 1 includes a recent guide²⁰ for assessing evidence on intervention effectiveness on three dimensions: the *strength* of evidence, which is determined by a combination of the study design (level), methodological quality and statistical precision; the *magnitude* of the measured effects; and *relevance* of the measured effects (as observed in the evaluation) to the implementation context. Such approaches are in tune with the epidemiological tradition of using multiple criteria to assess causal associations or causal inference (also listed in table 1). For the purpose of evaluating evidence on public health interventions, such an approach could be expanded to consider issues of intervention theory, intervention implementation, and monitoring in the evaluation process.

(c) The contribution of observational studies

Observational studies may represent the most feasible, acceptable and/or appropriate study designs for evaluating health interventions,⁵⁴ including public health interventions.⁵⁵ While RCTs (notably cluster RCTs) can be designed to evaluate even complex public health programmes,

often they are not feasible because of practical or resource constraints. Consequently well conducted RCTs are rare in public health. The implications of our reliance on observational evidence are threefold. We need to (a) better discriminate between different observational designs (b) improve our understanding of the bias in observational studies and (c) be pragmatic about the importance of study design relative to other dimensions of quality in evaluation research.

There are many useful observational designs available, including quasi-experimental designs, but guides for appraising evidence about clinical interventions do not discriminate among them. Their relative strengths and weaknesses are well described.^{8 10} The different study designs provide for alternative methods of assembling comparison groups, and of timing the implementation of an intervention in relation to the timing of various measurements. Thus, for example, replicated findings from interrupted time series designs (with repeated measurements before and after an intervention), by different investigators in different settings, may provide convincing evidence that an intervention is effective.

The relative validity of observational studies compared to RCTs has been the subject of much, ongoing debate among experts in evaluation methodology. Some studies have indicated that good observational designs can produce similar findings to those produced by RCTs, although "more empirical evidence is needed".⁵⁶ Conversely many of the comparative studies have themselves been critiqued for being methodologically flawed and highly confounded.⁵⁷ In an attempt to overcome such problems a recent study constructed randomised and non-randomised comparisons from a single dataset.⁵⁸ The authors concluded that non-randomised designs introduced "serious" and "unpredictable biases" that can lead to "both over- and under-estimates" of intervention effectiveness.

The potential for bias in observational studies will mean that their classification to lower *levels* (compared with RCTs) in the hierarchy of study design may be upheld. We do not seek to overturn such classifications and recognise that study design is highly important in evaluating evidence in public health. We do question however, the relative weight that is given to study design compared to other aspects of quality when appraising programme evaluations (as outlined in this paper).

Finally, appraisals of evidence quality are important in so far as they influence decisions about public health policy or practice. Care is needed that the use of evidence hierarchies to compare the potential for bias between study designs does not translate into unrealistic or overly expensive demands for level 1 or level 2 evidence, particularly if there is good or adequate level 3 evidence to inform a decision. In connection with this point, Kreuter aptly quoted Voltaire's aphorism in stating that "the best is the enemy of good" (Kreuter, 11th National Health Promotion Annual Conference, Perth, 1999).

WHAT ARE THE INTERVENTION OUTCOMES?

An evaluation of the adequacy of evidence about an intervention should include an examination of the range of outcomes considered. The evaluation criteria should help to determine whether the measured outcomes encompassed (a) the interests of people who might be involved in deciding on or delivering the intervention and (importantly) those affected by it; (b) unanticipated as well as anticipated effects of the intervention, beneficial or otherwise; and (c) the efficiency of the intervention, as well as its effectiveness.

(a) Identification of outcome information needed by important stakeholders

Given the social and political nature of public health, an appraisal of evidence should determine whether the outcome variables cover the interests of all the important stakeholders,

Table 2 Schemata for appraising economic evidence

Type of schema and evidence to be appraised	How the schema works and the criteria used
<p><i>Criteria for standardisation of methods and comparison of findings in economic evaluations</i> Appraise cost effectiveness studies, including cost utility evaluations.^{71 72}</p>	Defined a standard 'reference case' cost effectiveness analysis for comparison across all studies, which included consideration of societal perspective; definition and measurement of costs and outcomes; estimating effectiveness; incorporating time preference and uncertainty and presentation of results.
<p><i>Critical appraisal checklists for economic evaluation</i> Appraise economic evaluations in clinical setting.^{73 74}</p> <p>Appraise generic economic evaluations.^{75 76}</p> <p>Guide the appraisal and preparation of journal articles.^{77 78}</p>	<p>A guide for the clinical interpretation of published economic evaluations, based on 3 questions</p> <ul style="list-style-type: none"> ● Are the results valid? ● What were the results? ● Will the results help patients in particular context? <p>Most critical appraisal criteria focus on the following: specification of decision context and perspective; measurement of costs and outcomes; adjustments for timing and uncertainty; incremental analysis; presentation of results.</p>
<p><i>Regulatory guidelines</i> Appraise pharmacoeconomic submissions to governmental regulatory bodies.⁷⁹⁻⁸²</p>	Outlines recommended practice in submitting pharmacoeconomic evaluations to regulatory bodies for inclusion on subsidised pharmaceuticals list. Cover the perspective to adopt; definition and measurement of costs and outcomes; discounting and uncertainty; presentation of results to aid decision making.
<p><i>Ethical principles of good practice for economic evaluations</i> Appraise privately financed economic evaluations.⁸³</p>	Code of conduct to govern relationship between sponsors and researchers. Called for methods to be explicit.

not just those who conduct or appraise evaluative research.⁶⁰ Important stakeholders include those with responsibility for implementation decisions as well as those affected by the intervention. Some of the latter may be in disenfranchised groups, and it is not always clear whose interests have been (or should be) considered in evaluative research.^{61 62} This recommendation is in keeping with a long tradition in the social sciences known as *utilisation focused* evaluation.⁶³

Identification of the appropriate range of outcomes that should be included in a piece of evaluative research is one part of a pre-evaluation procedure known as "evaluability assessment".^{64 65} This was developed in the programme evaluation field more than two decades ago and has been popularised widely within health promotion.⁴⁴ Evaluability assessment requires a priori agreement about the successful outcomes of an intervention from important stakeholders' perspectives, including agreement on the types of evidence deemed to be adequate to reach a conclusion on the value of an intervention, and the questions to be asked in evaluating the intervention.⁶⁶

(b) Anticipated and unanticipated effects

Public health programmes often combine biomedical, educational, social and policy strategies that have many possible outcomes, such as changes in health states and determinants of health, processes, and characteristics of individuals, communities and environments.⁶⁷ These outcomes may be anticipated or unanticipated, and they may be intended or unintended. Unintended effects may be as desirable as, or more desirable than, the intended effects of the intervention. Conversely, unintended effects may detract from the intended effects to such an extent that assessment of the success of the intervention warrants revision. Evaluative research that records only the intended outcomes of an intervention may fail to detect its other positive or negative consequences. The methods of "goal-free evaluation"⁶⁸ are available for detecting unintended programme effects.⁶⁹

(c) Efficiency of interventions

Evidence-based health care is intended to take account of efficiency as well as effectiveness, although to date efficiency

questions have not been emphasised in evidence-based medicine.⁷⁰ The appraisal of evidence on public health interventions must inevitably determine whether efficiency has been assessed, and if so, how well. Examples of evidence-evaluation checklists that have been developed for appraising economic evaluations are listed in table 2. These include a guide to common standards so that evaluations from different settings can be compared; checklists for appraising published articles; regulatory guidelines; and ethical principles of good practice in economic evaluations.

IS THE RESEARCH EVIDENCE TRANSFERABLE?

Evidence-based decisions on the value and applicability of an intervention draw on knowledge of the effectiveness of an identical, similar or analogous intervention, usually carried out and evaluated in a different setting at a different time. To assess the transferability of evidence about an intervention information is needed on (a) the intervention itself (b) the evaluation context, and (c) interactions between the intervention and its context. A major limitation of traditional appraisal criteria is their inattention to adequacy of these aspects of the evidence.

(a) Information on the intervention

Public health interventions are rarely a standard package. To assess transferability, information is needed on the multiple components of an intervention. This should include details about the design, development and delivery of the various intervention strategies. Information is also needed on the characteristics of people for whom the intervention was effective, and the characteristics of those for whom it was less effective or even harmful. For many interventions, knowledge of factors that influence its sustainability and dissemination will also be important.^{44 67 84-87} These factors may be inherent to the way intervention strategies were delivered, or relate to the context in which they were implemented (see below). The availability of such information is a marker of the quality of evidence on public health interventions.

Table 3 Schemata for appraising qualitative or interpretive evidence

Type of schema and evidence to be appraised	How the schema works and the criteria used
<p><i>Critical appraisal checklists for determining the quality of qualitative research papers</i></p> <p>Several generic checklists have been published on how to appraise qualitative studies.^{101–104}</p>	<p>Evaluation schema considered for variety of qualitative studies. Quality is primarily determined by audit trail of research process and decisions made and credibility of study methods:</p> <ul style="list-style-type: none"> • Clarity of objectives and research questions • Appropriate selection of method to meet aims • Clear rationale for sampling strategy • Appropriate use of triangulation • Audit trail in data collection and analysis • Explicit researcher position and role • Clear basis for findings • Transferability of findings • Relevance, usefulness, importance of findings
<p><i>Standards to be used for a systematic review of qualitative literature</i></p> <p>Aimed at qualitative health services research.¹⁰⁰</p>	<p>Qualitative research is considered in the context of conducting systematic reviews on defined questions. Criteria for quality are intended to give priority to the:</p> <ul style="list-style-type: none"> • Interpretation of subjective meaning • Description of social context • Attention given to lay knowledge <p>Qualitative studies are assessed in following dimensions: theoretical basis of study; sampling strategy; scope of data collection; description of data collected; generalisability or typicality</p>
<p><i>Guide to appraising validity and applicability of qualitative literature</i></p> <p>One of the JAMA series: User's Guides to the Medical Literature, aimed at appraising qualitative research in clinical settings.¹⁰⁵</p>	<p>Examines four main aspects of qualitative analysis:</p> <ul style="list-style-type: none"> • Participant selection • Appropriateness of data collection • Process of data collection • Data analysis and corroboration of findings through triangulation

Table 4 Schemata for appraising process evaluations and evidence on intervention implementation (quality and quantity)

Type of schema and evidence to be appraised	How the schema works and the criteria used
<p><i>Assessments for health promotion evaluation studies</i></p> <p>International Union of Health Promotion guide on the appraisal of published reports on evaluations of health promotion and health education interventions and programmes.¹⁰⁶</p>	<p>Reviewers' assessments of the adequacy with which</p> <ul style="list-style-type: none"> • Intervention objectives are phrased (for example, degree of specification) • Situational framework is described (for example, how well the causal logic is described, the degree to which there are environmental barriers to the intervention, steps made to enlist the interest and cooperation of the program beneficiaries) • Intervention "correspondence" (for example, the extent to which the intervention is tailored to local circumstances, rewards or reinforces progress towards goals, goodness of fit between variables measured and programme objectives)
<p><i>Guides for appraising process evaluations of human services programs</i></p> <p>Assess programme quality, programme implementation and programme congruence, for health education programs and within organisational environments.^{67 85 107}</p>	<p>Recommendations provided on assessing evaluation methods and measures undertaken of the following: programme implementation; recipient response; site or host organisational response; practitioner response; personnel competencies and broader political and environmental factors affecting programme performance.</p>

(b) Information on the context

The social, organisational and political setting (or context) in which a public health intervention is implemented usually influences the intervention's effectiveness.^{85 88} It is important to distinguish between components of interventions that are highly context dependent (for example, a public education campaign to enhance immunisation uptake) and those that may be less so (for example, the efficacy of the vaccine itself among healthy infants). Contextual factors that influence the generalisability of evidence about interventions include literacy, income, cultural values and access to media and services.⁸⁹ Yet much published evidence on public health interventions does not include description of contextual variables or assess their impact on measures of effect.^{90 91} We should note that the lack of contextual information is also a weakness of evidence on medical interventions. For example, reports on surgical procedures often omit information on training, skill and experience of operators, or

even proxies such as hospital throughput. Established critical appraisal criteria do not draw attention to this deficiency.

(c) Information on interactions between the intervention and the context

Contextual factors often interact with interventions, even simple interventions such as educational programmes.⁹² Effect modification may arise from components of an intervention (for example, the skill and experience of the professional public health personnel responsible for the intervention), and/or the context (for example, cultural characteristics of the community in which the intervention was studied). Interactions between interventional and contextual components can have two implications. Firstly, they are likely to affect the transferability of the intervention and they also make an assessment of its transferability more difficult. Secondly, interactions greatly complicate attempts to pool the results of different

studies. Criteria for assessing evidence on public health interventions should therefore determine whether interactions have been sought, understood and explained. Where strong interactions are known to exist between an intervention and its context, it can be preferable (and more informative) to explore and explain their effects, rather than pooling the findings.

The information needed to assess the transferability of evidence is often drawn from research that uses a combination of different types methods, including observational, multilevel and qualitative methods (submitted data). Qualitative research can also enrich the understanding of intervention effects^{94–95} and guide systematic reviews.⁹⁶

Standards for conducting qualitative investigations are widely available.^{97–99} Recent interest in critical appraisal has stimulated the publication of several schemata for appraising qualitative research as a source of evidence.^{100–105} Examples of guides for evaluating qualitative evidence, and evidence from health promotion programme evaluations that focus on process and context information, are identified in tables 3 and 4.

EVIDENCE APPRAISAL AND PUBLIC HEALTH DECISIONS

Decisions about public health interventions should be based on a broad assessment of the strengths, weaknesses and gaps in the evidence. Reliance on levels of evidence alone to grade recommendations for action can attenuate public health decisions. For example, decisions that are mainly determined by criteria of evaluation study design will favour interventions with a medical rather than a social focus, those that target individuals rather than communities or populations, and those that focus on the influence of proximal rather than distal determinants of health.^{55 108 109}

It is also important to recognise the relative capacity of competing stakeholders (in a decision process) to generate evidence. Certain types of interventions (for example, pharmaceutical) are more likely to be supported by high quality evidence, simply because more resources are available to conduct the evaluation and produce that evidence (rather than because the interventions are better). In addition, “best” evidence is often gathered on simple interventions and from groups that are easy to reach in a population.¹⁰⁸ Thus conversely, little level 1 evidence exists on interventions for disadvantaged groups. This suggests that considerations of equity should temper the rigid application of rules of evidence in formulating recommendations for the use of public health resources.¹⁰⁸

Critical appraisal guides that identify and appraise multiple dimensions of evidence²⁰ permit greater scope for issues of relevance and transferability to be taken into account when formulating recommendations for practice. Yet it is still important to distinguish between a systematic and rigorous appraisal of available evidence, and the complex, socio-political process that determines policy and practice decisions. These distinct judgements will often be made by separate groups and are guided by different criteria and values.

Decisions about practice require a weighing of multiple factors such as the perceived magnitude and importance of the problem, the potential effectiveness and harms of the intervention, the feasibility of its implementation, its political acceptability, and the public demand for action. Different interest groups may advocate for competing recommendations,¹¹⁰ and recommendations based on the same evidence may change over time or change between contexts.^{111 112} In policy debate, a lack of good quality information about a problem can be interpreted as meaning that the problem is unimportant.¹¹³ As the notion of evidence-based policy gains substantial political currency,^{114 115} there is an analogous risk that a lack of high level evidence about the

Key points

- The evaluation of evidence about public health interventions should examine not only the credibility of the evidence, but also its completeness and its transferability
- The criteria used for critically appraising evidence need to reflect contemporary standards for planning and evaluating community-based programmes.
- The term “best quality” evidence should refer to evaluative research that was matched to the stage of development of the intervention; was able to detect important intervention effects; provided adequate process measures and contextual information, which are required for interpreting the findings; and addressed the needs of important stakeholders.

effectiveness of an intervention will exclude potentially valuable interventions from consideration. A clear distinction between criteria for evaluating evidence to determine what we know (and what we don’t know) about public health interventions, and the context dependent and often variable factors that determine local priorities, may allay such concerns.

CONCLUSION

The appraisal of evidence about public health interventions should encompass not only the credibility of evidence, but also its completeness and its transferability. The evaluation of an intervention’s effectiveness should be matched to the stage of development of that intervention. The evaluation should also be designed to detect all the important effects of the intervention, and to encapsulate the interests of all the important stakeholders.

These elements of evaluation have not yet been accepted as criteria for appraising evidence on public health interventions, although they are widely accepted in standards for planning and evaluating community-based programmes. We advocate their incorporation into criteria for appraising evidence on public health interventions. This can strengthen the value of evidence summaries and their potential contribution to the processes of public health advocacy and social development. Best quality evidence in public health is vital, but we should refrain from using the phrase “level 1” or “best” evidence synonymously with what is only one aspect of evidence quality, that is, study design.

Authors’ affiliations

L Rychetnik, M Frommer, Effective Healthcare Australia, School of Population Health and Health Services Research, University of Sydney, Australia

P Hawe, A Shiell, Department of Community Health Sciences, University of Calgary, Canada and School of Public Health, LaTrobe University, Australia

REFERENCES

- 1 **Oxman AD**, Sackett DL, Guyatt GH. Users guides to the medical literature I. How to get started. The evidence based medicine working group. *JAMA* 1993;**270**:2093–5.
- 2 **Woolf SH**, Battista RN, Anderson GM, *et al*. Canadian Task Force on the Periodic Health Examination—assessing the clinical effectiveness of preventive maneuvers: analytic principles and systematic methods in reviewing evidence and developing clinical practice recommendations. *J Clin Epidemiol* 1990;**43**:891–905.
- 3 **US Preventive Services Task Force**. *Guide to clinical preventive services*. 2nd edn. Baltimore: Williams and Wilkins, 1996.
- 4 **Guyatt GH**, Sackett DL, Cook DJ. Users’ guides to the medical literature II. How to use an article about therapy or prevention. A. Are the results of the study valid? *JAMA* 1993;**270**:2598–601.
- 5 **Guyatt GH**, Sackett DL, Cook DJ. Users’ guides to the medical literature II. How to use an article about therapy or prevention. B. What were the results and will they help me in caring for my patients? *JAMA* 1994;**271**:59–63.
- 6 **Hawe P**. How much trial and error should we tolerate in community trials? [Letter]. *BMJ* 2000;**320**:119.

- 7 **Suchman EA.** *Evaluative research.* New York: Russell Sage Foundation, 1967.
- 8 **Cook TD,** Campbell DT. *Quasi-experimentation; design and analysis issues for field settings.* Chicago: Rand McNally College Publishing, 1979.
- 9 **Borg WR,** Gall MD. *Educational research.* 4th edn. New York: Longman, 1983.
- 10 **Campbell DR,** Stanley JC. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally College Publishing, 1963.
- 11 **Canadian Task Force on the Periodic Health Examination.** The Periodic Health Examination. *Can Med Assoc J* 1979;**121**:1193–254.
- 12 **Sackett DL.** The Cochrane Collaboration. *ACP Journal Club* 1994;**120** (suppl 3):A-11.
- 13 **Centre for Evidence-Based Medicine.** *Levels of evidence.* <http://cebmlr2.ox.ac.uk/docs/levels.html>
- 14 **Sackett DL,** Richardson WS, Rosenberg W, et al. *Evidence-based medicine; how to practice and teach EBM.* New York: Churchill Livingstone, 1997.
- 15 **Public Health Resource Unit,** National Health Service and Institute of Health Sciences, Oxford. (<http://www.public-health.org.uk/casp/rct.html>)
- 16 **Thomas H,** Siracusa L, Ross G, et al. *Effectiveness of school-based interventions in reducing adolescent risk behaviour: a systematic review of reviews.* Prepared by the Effective Public Health Practice Project for the Public Health Branch. Ontario Ministry of Health, Canada, 1999.
- 17 **Mulrow CD,** Oxman AD, eds. *Cochrane Collaboration Reviewers Handbook.* In: *The Cochrane Library.* The Cochrane Collaboration. Oxford: Update Software; Issue 4, 1999 or online: <http://www.update-software.com/ccweb/cochrane/hbook.htm>
- 18 **The Campbell Collaboration.** <http://learning.gse.upenn.edu/campbell/>
- 19 **Liddle J,** Williamson M, Irwig L. *Method for evaluating research and guideline evidence.* Sydney: NSW Department of Health, 1996.
- 20 **National Health and Medical Research Council,** NHMRC. *How to use the evidence: assessment and application of scientific evidence.* Canberra: Commonwealth of Australia, AusInfo, 2000.
- 21 **Briss PA,** Zaza S, Pappaioanou M, et al. The Task Force on Community Preventive Services. *Am J Prev Med* 2000;**18** (suppl 1):35–43.
- 22 **NHS Research and Development Center for EBM.** <http://cebmlr2.ox.ac.uk/docs/levels.html>
- 23 **Hill AB.** *A short textbook of medical statistics.* 11th edn. London: Hodder and Stoughton, 1984.
- 24 **US Surgeon General's Advisory Committee on Smoking and Health.** *Smoking and health: report of the advisory committee to the Surgeon General of the Public Health Service.* US Dept of Health, Education and Welfare, Public Health Service. Washington, DC:US Government Printing Office, 1964.
- 25 **International Agency for Research on Cancer.** <http://193.51.164.11/monoeval/preamble.html>
- 26 **International Union for Health Promotion and Health Education.** *Health promotion program evaluation: Section B & C on estimation of intervention quality.* Copenhagen: 1994.
- 27 **Mosteller F,** Colditz GA. Understanding research synthesis (meta-analysis). *Annu Rev Public Health* 1996;**17**:1–23.
- 28 **Glasziou P,** Longbottom H. Evidence-based public health practice. *Australian and New Zealand Journal of Public Health* 1999;**23**:436–40.
- 29 **Carande-Kulis VG,** Maciosek MV, Briss PA, et al. Methods for systematic reviews of economic evaluations for the Guide to Community Preventive Services. *Am J Public Health* 2000;**18** (suppl 1):75–91.
- 30 **Kelly JG.** Ecological constraints on mental health services. *American Psychologist* 1996;**21**:531–9.
- 31 **McLeroy KR,** Bibeau D, Steckler A, et al. An ecological perspective on health promotion programs. *Health Educ Q* 1988;**15**:351–77.
- 32 **Winett RA,** King A, Altman D. *Health psychology and public health.* New York: Pergamon Press, 1989.
- 33 **Ziglio E.** How to move towards evidence-based health promotion interventions. *Promot Educ* 1997;**IV**:29–33.
- 34 **Baum F.** *The new public health. An Australian perspective.* Melbourne: Oxford University Press, 1998:127.
- 35 **Oakley A.** Experimentation and social interventions: a forgotten but important history. *BMJ* 1998;**317**:1239–42.
- 36 **Smith PJ,** Moffat MEK, Gelskey SC, et al. Are community health interventions evaluated appropriately? A review of six journals. *J Clin Epidemiol* 1997;**50**:137–46.
- 37 **Ebrahim S,** Davey Smith G. Effects of the Heartbeat Wales programme: effects of government policies on health behaviour must be studied. *BMJ* 1998;**317**:886.
- 38 **Harvey K.** Effects of the Heartbeat Wales Programme: health promotion is a waste of time and money. *BMJ* 1998;**317**:887.
- 39 **Shefer A,** Briss P, Rodewald L, et al. Improving immunisation coverage rates: an evidence-based review of the literature. *Epidemiol Rev* 1999;**21**:96–142.
- 40 **Susser M.** The tribulations of trials – intervention in communities. *American Journal of Health Promotion* 1995;**85**:156–60.
- 41 **Winkleby M.** The future of community-based cardiovascular disease intervention studies. *Am J Public Health* 1994;**84**:1369–72.
- 42 **Tudor-Smith C,** Nutbeam D, Moore L, et al. Effects of the Heartbeat Wales programme over five years on behavioural risks for cardiovascular disease: quasi-experimental comparison from Wales and a matched reference area. *BMJ* 1998;**316**:818–22.
- 43 **Sorensen G,** Emmons K, Hunt MK, et al. Implications of the results of community intervention trials. *Annu Rev Public Health* 1998;**19**:379–416.
- 44 **Hawe P,** Degeling D, Hall J. *Evaluating health promotion: a health workers guide.* Sydney: MacLennan and Petty, 1990.
- 45 **Campbell M,** Fitzpatrick R, Haines A, et al. Framework for design and evaluation of complex interventions to improve health. *BMJ* 2000;**321**:694–6.
- 46 **Flay B.** Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs. *Prev Med* 1986;**15**:451–74.
- 47 **Furberg CD.** Challenges to the funding of preventive research. *Prev Med* 1994;**23**:599–601.
- 48 **Krieger N,** Zierler, S. What explains the public's health? A call for epidemiologic theory. *Epidemiology* 1996;**7**:107–9.
- 49 **Krieger N,** Zierler, S. The need for epidemiologic theory. *Epidemiology* 1997;**8**:212–13.
- 50 **Diez-Roux AV.** Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;**88**:216–22.
- 51 **Susser M,** Susser, E. Choosing a future for epidemiology: I. Eras and paradigms. *Am J Public Health* 1996;**86**:668–73.
- 52 **Susser M,** Susser E. Choosing a future for epidemiology: II. From black box to Chinese boxes and eco-epidemiology. *Am J Public Health* 1996;**86**:674–7.
- 53 **Gyorkos T,** Tannenbaum TN, Abrahamowitz M, et al. An approach to the development of practice guidelines for community health interventions. *Can J Public Health* 1994;**85** (suppl 1):S8–13.
- 54 **Black N.** Why we need observational studies to evaluate the effectiveness of healthcare. *BMJ* 1995;**312**:1215–18.
- 55 **Mackenbach JP.** Public health epidemiology. *J Epidemiol Community Health* 1995;**49**:333–4.
- 56 **Ioannidis JP,** Haidich A, Lau J. Any causalities in the clash of randomised and observational evidence? No-recent comparisons have studied selected questions, but we do need more data. *BMJ* 2001;**322**:879–80.
- 57 **The Cochrane Collaboration Methods Groups Newsletter** 2001;**5**:16. (<http://www.cochrane.de/newslett/mgnews5.pdf>)
- 58 **Deeks JJ,** D'Amico, Sakarovich C, et al. How big are the biases associated with non-randomised designs used in evaluations of healthcare interventions? An empirical investigation. Abstract from paper oral presentation at the 3rd Oxford Symposium on Systematic Reviews, July 2000. (<http://www.ihs.ox.ac.uk/csm/Oralabstract2K.htm#Deeks1>)
- 59 Reference withdrawn.
- 60 **Lomas J.** Research and evidence-based decision making. *Aust N Z J Public Health* 1997;**21**:439–44.
- 61 **Kerridge I,** Lowe M, Henry D. Ethics and evidence based medicine. *BMJ* 1998;**316**:1151–3.
- 62 **Little M.** Assignments of meaning in epidemiology. *Soc Sci Medicine* 1998; **47**:1135–45.
- 63 **Patton MQ.** *Utilisation focussed evaluation.* 3rd ed. Thousand Oaks: Sage, 1997
- 64 **Wholey JS.** Evaluability assessment. In: Rutman L. *Evaluation research methods: a basic guide.* California: Sage Publications, 1977
- 65 **Scanlon JW,** Horst P, Jay JN, et al. Evaluability assessment: avoiding Type III and Type IV errors. In: Gilbert GR, Conklin PJ, eds. *Evaluation management: a sourcebook of readings.* Charlottesville: US Civil Service Commission, 1997.
- 66 **Rutman L.** *Evaluation research methods: a basic guide.* California: Sage Publications, 1977.
- 67 **Green LW,** Lewis FM. *Measurement and evaluation in health education and health promotion.* Palo Alto:Mayfield, 1986.
- 68 **Scriven M.** The pro and cons about goal free evaluation. *Evaluation Comment* 1972;**3**:1–7.
- 69 **Hawe P.** Capturing the meaning of 'community' in community intervention evaluation: some contributions from community psychology. *Health Promotion International* 1994;**9**:199–210.
- 70 **Maynard, A.** Evidence-based medicine: an incomplete method for informing treatment choices. *Lancet* 1997;**349**:126–8.
- 71 **Russell LB,** Gold MR, Siegel JE, et al. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996;**276**:1172–7.
- 72 **Weinstein MC,** Siegel JE, Gold MR, et al. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996;**276**:1253–8.
- 73 **Drummond MF,** Richardson WS, O'Brien BJ, et al. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? Evidence-Based Medicine Working Group. *JAMA* 1997;**277**:1552–7.
- 74 **O'Brien BJ,** Heyland D, Richardson WS, et al. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. B. What are the results and will they help me in caring for my patients? Evidence-Based Medicine Working Group. *JAMA* 1997;**277**:1802–6.
- 75 **Drummond MF,** Stoddart G, Torrance GW. *Methods for the economic evaluation of health care programmes.* Oxford: Oxford University Press, 1997.
- 76 **CASPInternational Online:** <http://www.phru.org/casp/economic.html>
- 77 **Drummond MF,** Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. *BMJ* 1996; **313**:275–83.
- 78 **Mason J,** Drummond MF. Reporting guidelines for economic studies. *Health Econ* 1995;**4**:85–94.
- 79 **Canadian Coordinating Office of Health Technology Assessment.** *Guidelines for economic evaluation of pharmaceuticals.* Ottawa: CCOHTA, 1997.
- 80 **Commonwealth of Australia.** *Guidelines for pharmaceutical industry on preparation of submission to the Pharmaceutical Benefits Advisory*

- Committee: including submissions involving economic analyses. Canberra: Australian Government Printing Office, 1992.
- 81 Department of Health guidelines for the economic evaluation of pharmaceuticals [press release 94/251]. London: Department of Health, 1994.
 - 82 **PhRMA Task Force on the Economic Evaluation of Pharmaceuticals.** *Methodological and conduct principles for pharmcoeconomic research.* Washington DC: Pharmaceutical Research and Manufacturers of America, 1995.
 - 83 **Task Force on Principles for Economic Analysis of Health Care Technology.** Economic analysis of health care technology. A report on principles. *Ann Intern Med* 1995;**122**:61–70.
 - 84 **Windsor RA,** Baranowski T, Clark N, *et al.* *Evaluation of health promotion and health education programs.* Palo Alto: Mayfield, 1984.
 - 85 **Roberts-Gray C,** Scheirer MA. Checking the congruence between a program and its organizational environment. In: Conrad KJ, Roberts-Gray C, eds. *Evaluating program environments. New directions in program evaluation.* No 40 Jossey Bass Higher Education and Social and Behavioral Sciences Series. San Francisco: 1988:63–82.
 - 86 **Durlak JA.** Why program implementation is important. *Journal of Prevention and Intervention in the Community* 1998;**17**:5–18.
 - 87 **Shediac-Rizkallah MC,** Bone LR. Planning for sustainability of community health programs: conceptual frameworks and future directions for research, practice and policy. *Health Educ Res* 1998;**13**:87–108.
 - 88 **Moos RH.** Understanding program environments. In: Conrad KJ, Roberts-Gray C, eds. *Evaluating program environments. New directions in program evaluation.* No 40 Jossey Bass Higher Education and Social and Behavioral Sciences Series. San Francisco: 1988:7–24.
 - 89 **Irwig L,** Zwarenstein M, Zwi A, *et al.* A flow diagram to facilitate selection of interventions and research for health care. *Bull World Health Organ* 1998;**76**:17–24.
 - 90 **Feinstein AR,** Horwitz RL. Problems in the 'evidence' of 'evidence-based medicine'. *Am J Med* 1997;**103**:529–35.
 - 91 **Birch S.** As a matter of fact: evidence-based decision-making unplugged. *Health Econ* 1997;**6**:547–59.
 - 92 **Watt D,** Verma S, Flynn L. Wellness programs: a review of the evidence. *Can Med Assoc J* 1998;**158**:224–30.
 - 93 Reference withdrawn.
 - 94 **Popay J,** Williams G. Qualitative research and evidence-based healthcare. *J R Soc Med* 1998;**91** (suppl 35):32–7.
 - 95 **Daly J,** MacDonald I. *Researching health care: designs, dilemmas, disciplines.* London: Tavistock/Routledge, 1992.
 - 96 **Oakley A.** Campbell Collaboration Website (2000). <http://campbell.gse.upenn.edu/papers.html>
 - 97 **Kirk J,** Miller MC. Reliability and validity in qualitative research. In: Maanen JR, ed. *Qualitative research methods series.* California: Sage, 1986.
 - 98 **Mays,** Pope C. Rigour and qualitative research. *BMJ* 1995;**311**:109–12.
 - 99 **Seale C,** Silverman D. Ensuring rigour in qualitative research. *European Journal of Public Health* 1997;**7**:379–84.
 - 100 **Popay J,** Rogers A, Williams G. Rationale and standards for the systematic review of qualitative literature in health services research. *Qualitative Health Research* 1998;**8**:341–51.
 - 101 **Greenhalgh T,** Taylor R. How to read a paper: papers that go beyond numbers (qualitative research). *BMJ* 1997;**315**:740–3.
 - 102 **Mays N,** Pope C. Assessing quality in qualitative research. *BMJ* 2000;**320**:50–2.
 - 103 **Usherwood T.** *Scientific quality criteria for qualitative research studies.* Sydney: Western Sydney Area Health Service. Scientific Advisory Committee, 2000.
 - 104 **Critical Appraisal Skills Program International Online:** <http://www.public-health.org.uk/casp/qualitative.html>
 - 105 **Giacomini MK,** Cook D. Users' guides to the medical literature: XXII. Qualitative research in health care; A Are the results valid?" *JAMA* 2000;**284**:357–62.
 - 106 **International Union for Health Promotion and Health Education.** Health promotion program evaluation: Section F on estimation of intervention quality. 1994.
 - 107 **Scheirer MA.** *Program implementation: the organisational context.* California: Sage, 1981.
 - 108 **Hawe P,** Shiell A. Preserving innovation under increasing accountability pressures: the health promotion investment portfolio approach. *Health Promotion Journal of Australia* 1995;**5**(2):4–9.
 - 109 **Institute of Medicine.** *Linking research and public health practice: a review of CDC's Program of Centers for Research and Demonstration of Health Promotion and Disease Prevention.* Washington DC: National Academy, 1997.
 - 110 **Sauerborn R,** Nitayarumphong S, Gerhardus A. Strategies to enhance the use of health systems for health sector reform. *Trop Med Int Health*, 1999;**4**:827–35.
 - 111 **Blakely,** T. Thornley, CN. Screening for hepatitis B carriers: evidence and policy development in New Zealand. *N Z Med J* 1999;**112**:431–3.
 - 112 **Woodward A,** Blakely T, Gander P, *et al.* Regulating public health in New Zealand. *Healthcare Review online* 1999;**3** or www.enigma.co.nz/hcro
 - 113 **Murray C,** Lopez AD. Evidence-based health policy lessons from global burden of disease studies. *Science* 1996;**274**:740–3.
 - 114 **Wooldridge M.** The importance of evidence-based medicine. Address by Honourable Dr Michael Wooldridge (Federal Health Minister). Opening of The Royal Australasian College of Physicians Annual Scientific Conference, Canberra. *Fellowship Affairs* 1996;September:29.
 - 115 **Downey M (reporter).** "Trust me, I'm a doctor". *Sydney Morning Herald* 10 May 1997.